PHYLOGENETICS: Fossil-free dating

Horizontal gene transfer events, the exchange of genetic material between organisms, can be used to date the timeline of evolution of microorganisms that lack a fossil record.

Mario dos Reis

Timing the main speciation events in the Tree of Life is one of the main goals of evolutionary biology. Biologists can achieve this goal by using molecular sequences to work out the relationships among groups of organisms, and then look at their fossil records to place the speciation events in the right geological context[1]. This task is, however, challenging because many organisms have scant, or even entirely absent, fossil records. Two contributions in this issue of *Nature Ecology and Evolution*, by Wolfe and Fournier[2], and Davín et al.[3], show how horizontal gene transfer (HGT) events can be used to date the Tree of Life of microorganisms that lack fossil records.

An HGT is the transfer of genetic material from the genome of a donor organism into the genome of a recipient. If the donor (or recipient) has a more complete fossil record, while the recipient (or donor) does not, then in principle it should be possible to use the HGT as a way to transfer the temporal information from one group into the other[3], and thus obtain a fully dated evolutionary tree for both groups (Fig. 1). Some HGTs are thought to have occurred several thousands of millions of years ago between the genomes of distantly related microorganisms, and thus the use of HGTs offers an exciting prospect for dating the most ancient events in the Tree of Life.

Wolfe and Fournier take advantage of an ancient gene transfer (that of SMC proteins[4]) from archaeal methanogens into Cyanobacteria to date the time of origination of methanogenesis (the biotic production of the gas methane). Cyanobacteria are among the most abundant photosynthetic bacteria on the planet, and they are important components of stromatolites, microbial mats formed in shallow waters by precipitation and/or trapping and binding of mineral sediments. Fossil stromatolites are the oldest fossil structures on the planet, some dated to 3.7 billion years ago[5] (Ga). On the other hand, archaeal methanogens have a scant fossil record, and this has precluded obtaining estimates of the origin of methanogenesis. Timing the origin of methanogenesis is important because it can inform the debate on the relative contributions of microbial versus abiotic sources of methane in the early history of the planet: Methane, a greenhouse gas, may have been present in high concentrations in the atmosphere of the young Earth, allowing liquid water to persist on the planet despite the weak radiation of a faint sun.

Wolfe and Fournier collected ribosomal and SMC protein sequences from several prokaryotic microorganisms (including Cyanobacteria and archaeal methanogens) for their analyses. A challenge when working with HGTs is that the tree depicting the evolutionary relationship among the species is different to the tree depicting the evolutionary relationships among the genes that have been horizontally transferred (Fig. 1). The authors overcame this challenge by 'splitting' the ribosomal protein alignment (which has evolved according to the species tree), so that it would not be in conflict with the SMC alignment (which involves the HGT and thus has a tree in conflict with the species tree). Wolfe and Fournier then used a Bayesian statistical approach that can appropriately accommodate patterns of protein evolution in distant groups[6], and that can integrate fossil constraints, to obtain a fully dated evolutionary tree of Cyanobacteria and methanogens. Their results provide a limit of no later than 3.51 Ga for the divergence of methanogens, with methanogenesis itself probably evolving earlier.

In the second work, Davín et al. performed an extensive analysis of HGT in the genomes of Archaea, Cyanobacteria and Fungi. Using a set of algorithms they developed previously, they calculate the relative ranking of species divergences that are compatible with the patterns of HGT observed. They then show that the ranking obtained correlates very well with the evolutionary distances obtained from the molecular data (in a similar way as the correlation seen between molecular distances and fossil ages according to the molecular clock hypothesis[7]). In other words, HGTs provide a temporal scaffold onto which we can place relative times of species divergences across the various groups. Davín et al. then show that evolutionary trees can be much better calibrated to geological time by incorporating the information from their scaffold of HGTs: timetrees that are incompatible with the HGTs can be removed from the analysis, thus improving the precision (that is reducing the uncertainty) in the estimates of the timing of diversification of the microbial groups.

Both works highlight how an ingenious toolkit, analysis of HGTs, can be used by evolutionary biologists in the quest to date the Tree of Life. The toolkit is, however, still at early stages of development. The 'splitting-of-alignment' technique used by Wolfe and Fournier to deal with species versus HGT tree conflict is impractical if the genes studied have been subjected to more than one HGT in the dataset, as this would involve splitting molecular alignments multiple times. Likewise, Davín et al. performed time estimation in two steps: They built Bayesian timetrees using genes with no HGT, and then used a separate analysis of HGTs to provide the relative temporal scaffold. Ideally, as Davin et al. suggest, an integrated Bayesian framework that deals with the tree conflict and corresponding temporal constrains should be developed. This should be feasible given the Bayesian method has become the statistical methodology of choice to date species divergences[8], from viral pandemics to the timescale of the Tree of Life.

Mario dos Reis

School of Chemical and Biological Sciences, Queen Mary University of London, London, WC1E 6BT, UK.
e-mail: m.dosreisbarros@qmul.ac.uk

References

1. Benton, M. J. & Donoghue, P. C. J. *Mol. Biol. Evol.* **24**, 26–53 (2006).
2. Wolfe, J. M. & Fournier, G. P. *Nat. Ecol. Evol.* (2018).
3. Davín, A. A., Tannier, E., Williams, T. A., Boussau, B., Daubin, V. & Szöllősi, G. J. *Nat. Ecol. Evol.* (2018).
4. Soppa, J. *Gene*. **278**, 253–264 (2001).
5. Nutman, A. P., Bennett, V. C., Friend, C. R. L., Van Kranendonk, M. J. & Chivas, A. R. *Nature*. 537, 535–538 (2016).
6. Lartillot, N. & Philippe, H. *Mol. Biol. Evol.* **21**, 1095–1109. (2004)
7. Zuckerkandl, E. & Pauling, L. in *Evolving Genes and Proteins* (eds Bryson, V. & Vogel, H. J.) 97–166 (Academic Press, 1965).
8. dos Reis, M., Donoghue, P. C. J. & Yang, Z. *Nat. Rev. Genet.* **17**, 71–80 (2016).

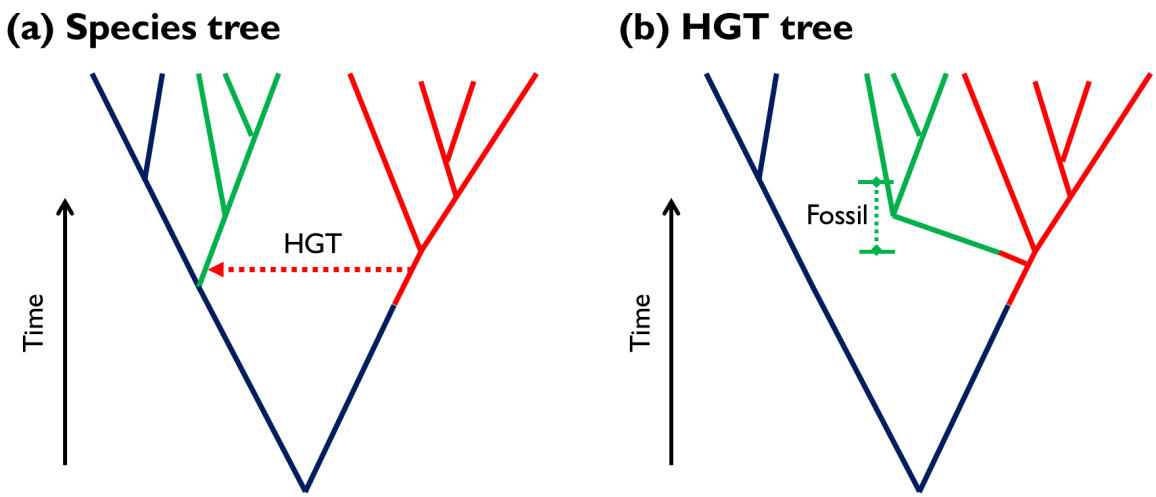

Figure 1. Horizontal gene transfers (HGT) can be used to date evolutionary trees for microorganisms that lack a fossil record. (a) A donor group (red) provides genetic material (dashed red arrow) to a recipient group (green). (b) An evolutionary tree estimated using the horizontally transferred gene will be different to the species tree. If a fossil constraint (green dashed line) can be assigned to a group (here the green group), then the HGT will be useful in transferring temporal information among groups and will allow the dating of the group lacking a fossil record (in this case the red group).